DOCUMENT RESUME

ED 153 009 CE 015 594

AUTHOR Wilson, Pamela A.; Wilson, Terry C.
TITLE What Factors Contribute to Better Instruction? Business Students Vs. Other Colleges.
PUB DATE [78]
NOTE 19p.; Paper presented at the annual meeting of the American Educational Research Association (Toronto, Canada, March 31, 1978)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Business Education; Colleges; College Students; College Teachers; *Comparative Analysis; Engineering; Evaluation; *Evaluation Criteria; *Evaluation Methods; Factor Analysis; Higher Education; Instruction; *Student Evaluation of Teacher Performance; Teacher Education

ABSTRACT

Student evaluations of instruction were factor analyzed to determine if students in the colleges of education, business, and engineering at Michigan State University varied in their perception of factors contributing to good instruction. The data base for the study consisted of evaluation forms that were processed during fall term, 1976, with a universe of 550 classes throughout the university. To produce stable results from factor analysis, approximately equal-size samples of 200 evaluations each were selected from the three colleges. Factor analyses were run on each individual college and on the combined sample containing all three colleges. In addition, separate analyses were run using the student, the class mean, and the student response removing the class mean as units of analysis. The results indicated that the variability in different factors differed considerably across colleges which may suggest that the evaluations require interpretations suitable to each college or that a unique form be developed for separate colleges. Additional research was supposed to obtain a large enough sample from each of the departments within the business school to recognize differences between departments and develop appropriate weighting schemes for each. (Author/VB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED153009

WHAT FACTORS CONTRIBUTE TO BETTER INSTRUCTION?

BUSINESS STUDENTS VS. OTHER COLLEGES

Pamela A. Wilson
Michigan State University

Terry C. Wilson
Wickes Corporation

Session # 33.02

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Pamela A. Wilson

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

OE 015 594

ABSTRACT

In an attempt to discover if students from the business college perceived factors contributing to good instruction differently than other colleges, student evaluations of instruction were factor analyzed. The schools included in this study were business, engineering and education. The results indicate that the variability in different factors differ considerably across colleges. This may suggest that the evaluations from business students require a different interpretation than evaluations from other schools, and possibly a unique form need be developed for separate colleges.

INTRODUCTION

In the past few years much time has been spent on the development of student evaluations of instruction. It has been noted by Baril and Skaggs (1975) that one of the major issues that should be addressed prior to data collection is the issue of whether different items or forms are necessary for different instructors, courses, departments or colleges within a university. This question becomes prompted by the underlying differences between courses in content, instructor emphasis, student characteristics and general academic discipline.

In a perusal of the student evaluation of instructor systems of the Big Ten Universities, it was found that different universities have varying opinions on Baril and Skaggs' (1975) issue. Several universities have differing instruments in use across the campus, however, if a department of evaluation services has developed an instrument it will be used by a large percentage of departments. Two examples of this are Michigan State University and University of Illinois at Urbana-Champaign. The Office of Evaluation Services at Michigan State University has built an evaluation instrument called the Level Two form that is administered to about 15,000 of the 44,000 students a quarter. While the University of Illinois administers the Illinois Course Evaluation Questionnaire (CEQ) to about 40% of the classes at UIUC.

On the other hand, Purdue University makes use of a cafeteria style rating system, i.e., the Purdue Rating Scale for Instruction. The Purdue scale is composed of both a common core section and an item pool of 300 behaviorally specific items. Each faculty member then has access to his own individualized evaluation instrument with a common core and twenty to thirty additional selected items.

It has been observed by the authors that a system like Purdue University is in a minority with regard to taking into consideration differences between instructors, courses, departments and colleges across universities. Student evaluation instruments are often developed and piloted on a very specific population, e.g., education, business, etc., as a sample of convenience. Once the instrument has been refined it is often used across colleges of the university. This has become increasingly true as teacher accountability has become a vital issue.

The question that the present study is concerned with is whether there are inherent differences in the development sample compared to the unsampled portion of the university. If differences do exist, do they invalidate the widespread use of one instrument across the university?

BACKGROUND

There has been developed in the field of education quite an extensive discourse concerning the reliability and validity of student evaluations during the past decade. In reviewing the literature on evaluations, many contradictions in research findings become evident.

Generally, student evaluation forms can be viewed as fulfilling three separate and independent purposes. Gillmore (1972) lists these three uses under the descriptive titles of normative, diagnostic, and informative. Others (Baril & Skaggs, 1975; Watruba & Wright, 1975) separate this trichotomy into user oriented terms; administrator, instructor and student information. 1) normative (administrator information) - evaluations used in a comparative mode as input for deciding salary increases, promotions, and other personnel evaluation considerations; 2) diagnostic (instructor information) - evaluations used to improve instruction; and 3) informative (student information) - evaluations used by students as an aid to class selection. Some schools have made attempts to implement three distinct instruments, but have often found resistance from students and faculty because of the extensive time and effort involved in completing three separate forms.

Given one or more forms, two issues that invariably arise are those pertaining to the reliability and validity of the instrument. A review of the evaluation literature with respect to the issues of reliability and validity, brings to light several aspects of each issue that is of concern to educators (Costin, Greenough & Menges, 1971).

First, the issue of reliability has produced several pertinent conclusions. Defining reliability as the extent to which an instrument consistently

measures a construct, it is commonly accepted that student evaluations achieve reliable results (Costin, Greenough & Menges, 1971; Mehrens, 1973). During the 1950's, the stability of students' ranking of the quality of teachers from year to year was established (Guthrie, 1976). Studies have consistently shown correlations from year to year to be in the vicinity of 0.90 (Costin, Greenough & Menges, 1971). Other issues that have been investigated include internal consistency of items within an instrument, intraclass correlation of ratings, and generalizability (dependability) ratings in lieu of classical reliability theory (Cronbach, et. al., 1972; Ebel, 1951; Kane, Gillmore & Crooks, 1976).

The second issue, that of the validity of student evaluation forms, is much less precise than the reliability issue. Generally defined, validity is the extent to which an instrument measures the construct it purports to measure. After such a general definition, validity is more precisely defined as face validity, content validity, and construct validity (Ebel, 1972). There are very encompassing definitions for each of these terms, but actual measurement of any type of validity is difficult at best. Conversely there is more and more demand for quantitative measurement of validity by students, faculty and administration.

Several studies attempting to quantify the validity measurements utilize correlations between the qualities an instrument purports to measure and certain explanatory constructs that at least partially explain performance on an instrument. Correlations between the following sets of variables have been shown useful: 1) supervisors and colleagues ratings (Costin, Greenough & Menges, 1971), 2) student ratings and teaching experience (Costin, Greenough & Menges, 1971), 3) student ratings and student GPA (Marsh, 1975), 4) student ratings and examination performance (Centra, 1977), 5) rating forms and student

achievement (Costin, Greenough & Menges, 1971; Fry, Leonard, & Beatty, 1975), and 6) student needs to teacher orientation (Tatenbaum, 1975). Unfortunately, the results of the studies noted above have not always been consistent, with some finding significant correlations and others finding the correlations to be statistically insignificant.

Recently, the issue of validity has taken a few new turns. Marsh (1977) has approached this topic by use of discriminant analysis. He has used data from graduating seniors who nominated instructors as "most outstanding" or "least outstanding" in conjunction with student evaluations from the following year. The validity of student evaluations of instruction was upheld in this study by the distinction between "most outstanding" and "least outstanding" being reflected in both data sources.

Another interesting look at validity is the "Dr. Fox Effect". One of the latest articles (Williams & Ware, 1977) on this phenomenon used an actor to deliver two lectures varying in substantive teaching points covered (high, medium, low) and expressiveness of the presentation (high, low). Unfortunately, the results observed to date suggest that student ratings of high expressive instructors may not reflect two important dimensions of teaching effectiveness, namely; substantiveness of instruction and degree of student achievement.

One area that has to do with both the reliability and validity of evaluation forms that researchers have not addressed to a great degree is the differences between colleges within a university. More than likely, the lack of published research on the topic of differences between colleges is somewhat politically motivated. After all, who is going to state that the professors from one college are inferior to those of another college in terms of teaching skills? However, there is a logical framework for hypothesizing that students

enrolled in different colleges use different criteria for evaluation. For example, the students in the physical sciences would probably use more cognitively oriented criteria, while those from the social sciences could use more affective criteria. Such differences could account for the conflicting results found in much of the research to date.

One problem that arises in dealing with the differences between colleges within a university is what statistical technique would be appropriate to perceive this difference. Traditionally, factor analysis programs have been run on student evaluation of instruction data to indicate general factors in the instrument. One possibility of comparing colleges would be to run individual factor analysis on colleges within a university and compare what items comprised predominant factors and the percent of variance accounted for by different factors.

In the past, the student response has been used as the unit of analysis in factor analysis. However, recent research (Doyle & Whitely, 1974; Linn, Centra & Tucker, 1975; Whitely & Doyle, 1976) has questioned this as the appropriate unit of analysis. It has been noted that there are three possibilities contending for the unit of analyses: 1) Student response - this technique would use correlations among items across students. This would confound sources of covariation between instructors (or classrooms) and the within classroom source of covariation. 2) The classroom mean on each item - in this situation a correlation among item means across classrooms would be obtained. A disadvantage here is that the number of observations is too small compared to the number of items. This would, however present a stable view of the instructor. 3) Within classroom deviations from classroom means - in this situation the analyses is revealing individual variability and consistency

across items, but not instructor variability.

METHODOLOGY

In this study, the same student rating form is used for all colleges within the university. Because the individual questions on the form had been previously factor analyzed in 1971 and 1976 (Office of Evaluation Services, 1971), it is reasonable to expect stability of factors for the present study. However, it is hypothesized that the amount of variance attributable to a particular factor will differ from school to school. Differences in variance attributable to a factor between schools would at least indicate that the differences be recognized when interpreting the evaluations.

The data base for the present study consists of those evaluation forms that were processed during Fall term, 1976. The universe consisted of 550 classes throughout the university. Three colleges within the university were chosen for analysis: 1) business, 2) education, and 3) engineering. These particular classes were chosen as they were hypothesized to attract different types of students. For example, education students would be more affective than engineering students, and business students should place somewhere between the two other schools on an affective scale. Classes were selected from each of the three colleges in an attempt to achieve an approximately equal sample size for each college. Fifteen classes were selected from Education, 19 from Business and 28 from Engineering, resulting in a sample size of more than 200 evaluations from each. A sample of this magnitude is required to produce stable results from factor analysis.

Factor analyses were run on each school individually and the combined sample containing all schools. Separate analyses were also run using the student, the class mean, and the student response removing the class mean as units of analysis. All computer analyses were done with the SPSS program, with both varimax and oblique rotation (Child, 1973; Kerlinger, 1973; Nie, Hull, Jenkins, Steinbrenner & Bent, 197 ..

ANALYSIS

Twenty four separate factor analysis runs were made. All computer analysis was done with the SPSS program. The twenty four analyses are outlined in Table 1. Table 2 presents a list of the variables used as input to the factor analysis program. All of the attitude variables required responses on a five point Likert scale: 1 = superior, 2 = above average, 3 = average, 4 = below average, 5 = inferior. Table 3 presents the factor names. The variable clusters differ somewhat between schools, but in general the following groups of variables could conceivably be referred to by the factor names indicated in Table 3 (Office of Evaluation Services, 1971). Of course, the factor names are subjectively specified, but they appear to be consistent with the intent of the variable clusters.

An examination of Tables 4 through 7 reveals the factors to be relatively homogeneous across schools. Only three factors with eigenvalues greater than one emerged from each of the factor analyses. The interesting and important aspect is that while the factor clusters are relatively consistent the percent of variance attributable to a particular factor for each school is very different. For example, the orthogonal analysis using scores produces different factors accounting for the majority of variance in each school. There is no

TABLE 1

FACTOR ANALYSIS RUNS

Data	For 4 colleges
Raw Scores:	
Orthogonal	X
Oblique	X
Z Scores:	
Orthogonal	X
Oblique	X
Variable Means:	
Orthogonal	X
Oblique	X
TOTAL	24

TABLE 2

LIST OF VARIABLES

Variable Number	Description
1	Instructor's enthusiasm
2	Instructor's interest in teaching
3	Instructor's use of examples
4	Instructor's concern with students' learning
5	Student interest in course
6	Student attentiveness in class
7	Course was an intellectual challenge
8	Student improved competence in course area
9	Student encouraged to express opinions
10	Instructor's receptiveness to new ideas and opinions
11	Opportunity to ask questions
12	Instructor's stimulation of class discussion
13	Appropriate amount of material covered
14	Course taught at appropriate pace
15	Contribution of homework assignments to understanding
16	Appropriateness of difficulty of assigned reading
17	Instructor's ability to relate concepts in a systematic manner
18	Course organization
19	Ease of note taking
20	Adequacy of course outline
21	General enjoyment of course
22	Was course required (yes/no)
23	Sex (M/F)
24	Overall GPA (a) 1.9 or less, (b) 2.0-2.2, (c) 2.3-2.7, (d) 2.8-3.3, (e) 3.4-4.0
25	Class level (F/So/J/S/G)

TABLE 4

ORTHOGONAL FACTORS BY SCHOOL*

		BUSINESS N=851 classes	ED N=606 classes	ENG N=834 classes	COMB. N=2291 classes
Factor 1:	Variables	9,10,11,	4,9,10,11,12,14	9,10,11,12	9,10,11,12
	Eigenvalue	9.2	10.4	8.5	9.4
	% of Variance**	67.3	69.8	65.7	68.7
2		1,2,3,4,17,21	5,6,7,8,21	5,6,7,8	1,2,3,4
		1.4	1.6	1.7	1.3
		10.6	10.9	12.9	9.7
3		13,14	13,17,18,19,20	13,14,15	5,6,7,8,21
		1.0	0.90	1.0	0.92
		7.3	6.0	8.0	6.8
4		5,6,7,8,21	1,2,3,4	1,2,3,4	13,14,15,16
		0.82	0.81	0.91	0.84
		6.0	5.4	7.0	6.2
5		17,18,19,20	13,14	17,18,19,20	17,18,19,20
		0.68	0.68	0.51	0.62
		5.0	4.6	3.9	4.5

*Variables in each column are variable numbers from Table 2.

**Percents do not sum to 100 because program was run with six factors but only the first five appeared as common factors.

TABLE 5

ORTHOGONAL FACTORS USING Z SCORES

	Business	Education	Engineering	Combined
Factor 1	9, 10, 11, 12	17, 18, 19, 20	5, 6, 7, 8, 21	9, 10, 11, 12
2	17, 18, 19, 20	9, 10, 11, 12	9, 10, 11, 12	5, 6, 7, 8, 21
3	5, 6, 7, 8, 21	5, 6, 7, 8, 15, 16, 21	1, 2, 3, 4	13, 14
4	1, 2, 3, 4	1, 2, 3, 4	13, 14, 15	17, 18, 19, 20
5	13, 14	13, 14	17, 18, 19, 20	1, 2, 3, 4

TABLE 6

OBLIQUE FACTORS BY SCHOOL*

	Business	Education	Engineering	Combined
Factor 1	17, 18, 19, 20	1, 2, 17, 18, 19, 20	1, 2, 3, 4	13, 14
2	9, 10, 11, 12	24, 25	9, 10, 11, 12	9, 10, 11, 12
3	1, 2, 3	5, 6, 7, 8, 21	5, 6, 7, 8, 21	1, 2, 3
4	25	9, 10, 11, 12	13, 14, 15	5, 6, 7, 8
5	13, 14, 15, 16	13, 14	17, 18, 19, 20	24, 25

*Oblique analyses produced consistent factors for delta values of -1.0, -0.5 and 0.0.

TABLE 7

OBLIQUE FACTORS USING Z SCORES*

	Business	Education	Engineering	Combined
Factor 1	17, 18, 19, 20	17, 18, 19, 20	17, 18, 19, 20	17, 18, 19, 20
2	9, 10, 11, 12	9, 10, 11, 12	9, 10, 11, 12	9, 10, 11, 12
3	1, 2, 3	5, 6, 7, 8	5, 6, 7, 8	1, 2, 3, 4
4	5, 6, 7, 8	1, 2, 3, 4	13, 14, 15	5, 6, 7, 8, 21
5	13, 14	24, 25	1, 2, 3, 4	13, 14

*(Same FN as Table 6)

First, it is apparent the teachers from different schools cannot validly compare their ratings without a differential weighting scheme. Second, self-evaluation must be considered in light of the fact that students apparently place more importance on some factors than others. Each teacher must compare his own perception of factor importance with that of the students, and decide in terms of specific courses the rank order that is appropriate. For example, many universities have classes of more than 100 students where student-instructor interaction is virtually non-existent. Third, personnel committees should recognize the implications of this study. The rank order of factors being different across schools, each schools' personnel committee must consider the weighting of the factors. Again, the rank orders must be appropriately defined in terms of specific courses. In the event that a school has all classes where each of the factors is equally appropriate for all teachers, the factors must be equally weighted before a relative ranking of teachers is possible.

Most universities have evaluation or education departments that could be an invaluable resource in developing appropriate evaluation schemes, or separate forms for each college. In any event, it is time that everyone teaching or evaluating at the college level recognize the structure of the evaluation forms they are using. The next phase of the present research is to obtain a large enough sample from each of the departments within the business school to recognize differences between departments and develop appropriate weighting schemes for each.

REFERENCES

Baril, G. L.; Skaggs, C. T. "Selecting Items for a College Course Evaluation Form," College Student Journal, Vol. 10, Summer, 1976, pp. 183-187.

Centra, J. A. "Student Ratings of Instruction and Their Relationship to Student Learning," American Educational Research Journal, Vol. 14, No. 1, pp. 17-24.

Child, D. The Essentials of Factor Analysis, New York: Holt, Rinehart and Winston, 1973.

Costin, F; Greenough, W. T.; Menges, R. "Student Ratings of College Teaching: Reliability, Validity, and Usefulness," Review of Educational Research, Vol. 41, No. 5, December 1971, pp. 511-535.

Cronbach, L. J.; Gleser, G. C.; Nanda, H.; Rajaratnam, N. The Dependability of Behavioral Measurements: Theory of Generalizability for Scores and Profiles, New York: John Wiley and Sons, Inc., 1972.

Doyle, K. O.; Whitely, S. E. "Student Ratings as Criteria for Effective Teaching," American Educational Research Journal, Vol. 11, No. 3, Summer 1974, pp. 259-274.

Ebel, R. L. Essentials of Educational Measurement, New Jersey: Prentice-Hall, Inc., 1972.

Ebel, R. L. "Estimation of the Reliability of Ratings," Psychometrika, Vol. 16, 1951, pp. 407-424.

Fry, P. W.; Leonard, D. W.; Beatty, W. W. "Student Ratings of Instruction: Validation Research," American Educational Research Journal, Vol. 12, No. 4, Fall 1975, pp. 435-447.

Gillmore, G. M. "Three Functions of Student Course Evaluations." Paper presented at a Symposium on Course Evaluation, NCME Convention, 1972.

Guthrie, E. R. The Evaluation of Teacher Teaching: A Progress Report. Seattle: University of Washington, 1954.

Kane, M. T.; Gillmore, G. M.; Crooks, T. J. "Student Evaluations of Teaching: The Generalizability of Class Means," Journal of Educational Measurement, Vol. 13, No. 3, Fall 1976.

Kerlinger, F. N. Foundations of Behavioral Research, New York: Holt, Rinehart and Winston, Inc., 1973, Chapter 37.

Linn, R. L.; Centra, J. A.; Tucker, L. "Between, Within, and Total Group Factor Analyses of Student Rating of Instruction," Multivariate Behavioral Research, July, 1975, pp. 277-288.